

Liz Levy[Help](#)Basic  
SearchAdvanced  
SearchTopic  
GuidePublication  
Search

Marked List

0 articles

Interface language:

English

Databases selected: Multiple databases...

**Article View**[<< Back to Results](#)[< Previous](#) Article 14 of 206 [Next >](#)[Publisher Information](#)[Print](#)[Email](#)☐ Mark Article[Citation](#) , [Full Text](#) , [Text+Graphics](#) , [Page Image - PDF](#)**Content management breaks down your docs**Liz Levy. Imaging & Document Solutions. San Francisco: Dec 1998. Vol. 7, Iss. 12; pg. 16, 8 pgs[» Jump to full text](#) Author(s): Liz LevyPublication title: Imaging & Document Solutions. San Francisco: Dec 1998. Vol. 7, Iss. 12; pg. 16, 8 pgs

Source Type: Periodical

ISSN/ISBN: 15226085

ProQuest document ID: 52044919

Text Word Count 3693

Article URL: [http://gateway.proquest.com/openurl?ctx\\_ver=z39.88-2003&res\\_id=xri:pqd&rft\\_val\\_fmt=ori:fmt:kev:mtx:journal&genre=article&rft\\_id=xri:pqd:did=00000005](http://gateway.proquest.com/openurl?ctx_ver=z39.88-2003&res_id=xri:pqd&rft_val_fmt=ori:fmt:kev:mtx:journal&genre=article&rft_id=xri:pqd:did=00000005)**Full Text** (3693 words)*Copyright Miller Freeman Inc. Dec 1998***[Headnote]**

Organizations are struggling with just how to bring their existing information online. Content management offers a powerful way to separate and recombine kernels of information for use on the Web, CD-ROM or any medium.

Where traditional document management ends, content management begins. Document management handles information from a file perspective. Each file is its own entity, and it is indexed, stored, retrieved and used at the file level. Content management breaks down a file into its component parts, and these elements are indexed, stored, retrieved and used at the content level.

A document file can be made up of multiple types of information, such as charts, tables, headlines, captions, text and even sound or video. Content management breaks out the content from the document. The information can then be accessed individually and brought together in different ways. As a result, you can "repurpose" information for different mediums, including print, CD-ROM and the Web.

With the emergence of the Web, many businesses are discovering that they are in the publishing business. According to analyst firm International Data Corp. of Framingham, MA, "One of the biggest challenges facing organizations deploying intranet applications is the ongoing organization and management of the dynamic set of interrelated documents [content], generally authored as a collaborative process, that compose the content found on these corporate sites."

Document management systems have served many organizations well in this regard, providing a way to share information internally and externally across intranets. But what document management has been lacking is the ability to use and reuse bits and pieces of information for different purposes and with many different delivery options. This is where content management comes in.

There is some overlap in the document management and content management (as well as knowledge

management) markets. "The process is more important than the product," says Priscilla Emery, senior VP of information products and services at AIIM International. "The product must accomplish your process. With the growth of the Web in business, everyone is becoming a publisher and content management deals with those publishing issues. It is a publishing process that works in front of the business process."

#### XML Provides a Foundation

XML (Extensible Markup Language) is a technology that is bringing together the separate worlds of document management and content delivery. Using XML, a document can be separated into its discrete elements. For example, a 50-page report might contain a five-page executive summary, three 15-page chapters, 20 charts, five tables, eight photographs, 40 headlines and 30 captions. XML offers a way to identify the different types of information presented and the relationships between that information.

We are all familiar with HTML (Hypertext Markup Language), the language most popularly used to display information on the Web. HTML deals primarily with presentation of electronic documents and does not offer much in the way of providing structure for the content. In the 5(h page report example used above, HTML would not be able to provide information on the difference or relationship between a headline and a caption, with the possible exception that they might be presented in bold type.

HTML was developed from the more complex SGML (Standard Generalized Markup Language), which was developed in 1978 with the idea that a standard "markup" could create structure separate from the appearance of the document. Markup is simply providing information along with a document that says something about its appearance or structure. SGML is a language that is used to code nearly any type of document to describe its structure. SGML is not as widely used as HTML because of its complexity and the high cost of SGML editing tools.

XML was developed as a simpler version of SGML XML documents are composed of many entities better described as objects, such as the Word, PowerPoint, TIF or Wave files that might make up a document. Each object can contain one or more elements. An element is a single bit of information usually encapsulated as a string of text. Each of these elements can have certain attributes or properties that will describe the way in which the information is to be processed.

XML provides a way of describing the relationships between these entities, elements and attributes. It tells the computer how it can recognize the component parts of a document. It accomplishes this through the use of metadata, which is simply coding information about the information. Metadata is any information that accompanies an XML document, including tags, DTDs, links and style sheets.

Elements in the document are marked with and identified by tags. The tags are placed at the beginning and end of a string of text to describe the attributes of the element. Tags do not describe the format whether the text is bold or italic. Nor do they instruct the computer what to do in a Web page when people click on it. Tags simply identify what the element is. This is different from HTML, where the tags must perform all these functions at once.

Enlarge 200%
Enlarge 400%



Separating the formatting from the element's identity makes it easier to repurpose information for different media. With HTML you have to write new tags each time you reuse an element for display on a different medium. With XML a Document Type Definition (DTD) lets you define the tags that were created and then use and reuse the elements.

A DTD is created once to describe the document type (i.e. expense report, fiscal statement, press release), what the content elements are and their structure. It does not define aspects of format, such as bold or centered. Formatting is done later through XML style sheets using a format not yet finalized called XSL (Extensible Style Language).

Once a file is created and submitted to the central repository, it is validated against the DTD and parsed, meaning its content elements are broken out. Validation is a key function of XML. It lets the structure of data be checked before it is used in an application.

Once parsed, an XML document is manipulated through an object model (or API). A standard DOM (Document Object Model) is being developed by the World Wide Web Consortium (WC3), the international body that develops standards for the Web.

The DOM lays out an XML document and its elements in a tree structure.

Simply stated, XML provides a standard file format for presenting data and a standard way to include information about the data (metadata) to describe its own structure. XML Version 1.0 was finalized as a W3C Recommendation in February of this year. Work is still underway on the accompanying style language, called XSL (Extensible Style Language), and a link language, called XLL (Extensible Linking Language).

The XML standard has been adopted by both Microsoft and Sun. Microsoft has implemented support for XML within Internet Explorer 4.0 and will continue support with 5.0. XML will also be adopted in the next release of Office and other Microsoft products. When these products arrive, you will be able to "Save as XML" as simply as you do any other format today. Standard DTDs are being developed for many other popular applications, but as of now, users will need to design their own DTDs, which is not an easy task.

## How Content

### Management Works

The model for content management is to store and manage XML content from a central repository, usually an object-oriented or relational database or a hybrid of both. Content management systems offer many of the same functions as document management systems, such as storage, version control, checkin/check-out and text searches. What's added is the ability to tag, index, search and reuse smaller elements of a document.

Content management systems output content from the central repository through any means of delivery available. Web delivery is particularly efficient. Take as an example a company that manufactures custom PCs that needs to provide customized manuals on demand for each configuration ordered. Content management would let this manufacturer break out and store content on individual PC parts as separate items with their relationships to other parts.

The content on just those parts ordered by a specific customer could be pulled together to create a manual. The manual could be printed, output to CD-ROM or served up on a Web site at the customer's request. The result is more efficient, dynamic and responsive use of content, and this translates into cost savings for the manufacturer and better, more personalized service for the customer.

Like many of today's document management systems, XML will take advantage of three-tier architectures that support Web applications. In this model, a browser front end interacts with a middle-tier Web server, which in turn communicates with a backend database server for central storage. Information can be converted to XML on the middle tier, offering new ways to access stored information from mainframes and databases. This data can then be delivered to the Web and exchanged online as easily as HTML pages using HTTP.

### A Look At Three Products

The three content management systems described below are providing innovative solutions for using information efficiently and for automating the publication of information to the Web. The main features that are common to all these products are a central repository based on a database, XML content, HTML conversion and Web publishing. Many of these vendors and products also have a history of supporting SGML. This may offer those looking to move from SGML to XML some added expertise and convenience.

DynaBase v3.0 from Ipso (Boston, MA 617-753-6500 [www.inso.com](http://www.inso.com)) is an XMLbased content management and web publishing system. XML content management capabilities include import, parse and store XML components with indexing and version control. Content for publishing can also be served on the Web with XML tag-level scripting and XML tag-level search and retrieval.

The content management components include a Data Server and a Web Manager Client. The Data Server plugs into existing Web servers for content management. It is a repository that includes a built-in object-oriented database for storing any and all Web content. The content can be XML, HTML, graphics, video, applets or scripts.

The Web Manager Client is the interface to the Data Server. From the Web Manager Client, users can browse for files and check out content from the Data Server or import new content using HTTP. The Web Manager automatically launches authoring applications when you want to edit content.

Inso's publishing system includes the Web Server Plug-In and the Web Developer Client. The system lets publishers create Web sites and pages on the fly.

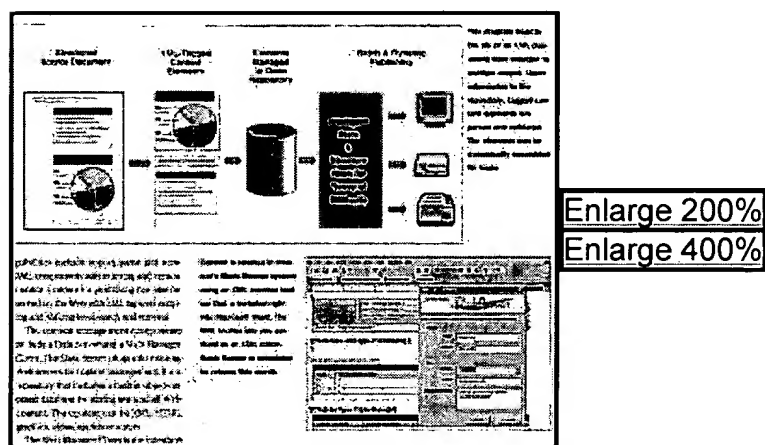
Version 3.0 of DynaBase, which was released in September, includes Inso's "Outside In" Web server-based file format converter, which automatically converts popular file formats like Microsoft Word, Excel and PowerPoint to HTML, GIF or JPEG. Also included are new workflow capabilities that can be applied to any authoring, editing and publishing process.

DynaBase supports Mac and Unix platforms through a cross-platform Java client. There is also added plug-in support for Microsoft Internet Information Server 4.0 running Windows NT and Netscape Enterprise Server 3.5.1 running Windows NT4.0 and Sun Solaris 2.6.

Starting pricing for a DynaBase enterprise installation is about \$47,000. This includes ten client licenses, one data server that supports unlimited users, three Web server plug-ins for management and delivery to the Web and ten host or domain names for supporting ten separate Web projects at once.

Blade Runner from Interleaf (Waltham, MA 781-29(?-1710 [www.interleaf.com](http://www.interleaf.com)) is an enterprise content management system scheduled for release this month. BladeRunner addresses the life cycle of content in three layers: content creation, content repository and content publish. Content is created in Blade Runner using an XML creation toolbar that is installed right into Microsoft Word. The XML toolbar lets you use Word as an XML editor. Standard DTDs are implemented and used to create Word templates that enforce certain rules or restraints on the author. Users do have to know XML. They create Word files using a template that matches the XML DTD. The author is guided with lists of choices for valid content. The document is then validated against the DTD and saved as an XML document. The document goes through a final validation before check in to the content repository.

A DTD creation tool, Microstar's Near and Far Designer, is included, but any other DTD creation tools can be used. Authors are given lists of choices for content structure that adhere to the XML DTD. The content is validated against the DTD before it is checked into the content repository.



This diagram depicts the life of an XML document from creation to multiple output. Upon submission to the repository, tagged content elements are parsed and validated. The elements can be dynamically assembled for reuse. Content is created in Interleafs Blade Runner system using an XML creation tool bar that is installed right into Microsoft Word. The XML toolbar lets you use Word as an XML editor. Blade Runner is scheduled for release this month.

Blade Runner's content repository is built on an object-oriented database from Poet (San Mateo, CA 650 281-9640 [www.poet.com](http://www.poet.com)). Once the content is validated, the XML document is "burst" (i.e., broken down) into its discrete elements. The reusable content elements can be brought together and formatted for output using the menu-driven Composer/Style Editor.

The database stores XML content along with the DTDs, XSL style sheets and XLinks. The system provides versioning, check-in, check-out, searching, navigational linking and referencing. Searches can be performed on text as well as content, properties, structure and metadata.

A tool is provided for creating XSL style sheets, which are applied to the XML content to render a document. Users can preview the layout and edit the final document.

The XML content is published using a two-part Batch Publishing Engine that prepares content for print, CD or the Web. Since browsers other than MS Internet Explorer have yet to natively support XML, the publishing engine converts content to HTML for presentation on the Web. The Assembly Engine uses Cascading style Sheets for batch conversion to HTML for the Web.


The XML toolbar can also be installed in the Interleaf 7 publishing system. Interleaf supplies data adapters with Blade Runner that support the integration of applications like SAP and Baan and the use of this information by the content management system as XML content. Blade Runner operates on Windows NT servers and Windows NT/9X clients.

Pricing had not been set yet but will be announced this month at the release date

announced this month at the release date and will be in the six figure range for enterprise applications, depending on how the solution is packaged.

Information Manager 2.0 from Texcel (Cambridge, MA 617-621-7004 [www.texcel.no](http://www.texcel.no)) puts a great deal of emphasis on collaboration. It is a content and "process" management system that includes collaborative tools for using content. Users can find, edit, review, reuse and assemble information managed from an Information Manager (IM) database.

IM identifies all the elements of a document defined by XML or SGML markup as separate objects. These can be searched, retrieved and accessed for editing or viewing. IM tracks object versions in the database and the software manages the Links between objects and documents along with metadata.

The system integrates with SGML/XML editors such as Adobe's FrameMaker+SGML and  ArborText's Adept Editor, making it easy to update XML and SGML data from the database in their native application. For Microsoft Word files and other types of files the application can be selected. Integrated ActiveX controls can be used to make different file types like MS Word files accessible from the database.

Using a Windows Explorer-like interface, users can browse multiple databases and click on an XML or SGML document. IM dig plays a view of a document's contents in a small window. Search tools let users find any content, link or metadata.

To support collaboration, an Electronic Review tool lets users insert comments that can be passed along with the document. IM includes a workflow system that lets you design and implement work processes.

IM's Document Assembly tool lets you gather content to create documents or publications that can be changed or updated for different purposes. A Document Assembly template can incorporate boilerplate text and database queries.

Visual Basic can be used to add customizations. Open APts are also available to developers. IM supports Windows 95 and Windows NT clients. It also can support Web browsers, but the separate IM Web Application is needed. Windows NT servers and Unix servers are supported.


Information Manager costs \$25,000 for the starter kit, which includes the server, software, development, workflow and four concurrent user licenses.

#### Document Managers Tackle Content

Many document management vendors have been slow to adopt XML. Those that have adopted it offer useful systems for managing information at the content level and delivering content to the Web as a front end to their enterprise systems.

Without XML, you cannot achieve the same level of granular information management. The 50-page report mentioned earlier could not be broken down into its discrete parts unless those parts were saved as separate files. To reuse these elements, you would have to cut and paste the content together.

This is the crucial difference between content management and compound document management, which has long been supported by many document management systems. With compound document management, different objects, such as Word files, Excel files, images, graphics, sound or video can be accessed from one system and pulled together for delivery to the Web.

 FileNet (Costa Mesa, CA 71496&3400 [www.filenet.com](http://www.filenet.com)) is taking on Web publishing and talking about "content" management with the introduction of Panagon Web Publisher, which was introduced in November. The product offers an automated solution for managing and incorporating multiple types of information and the delivery of that information to the Web.

Panagon Web Publisher distributes the job of the typical, overburdened Web Master by providing tools to automate the publication of information to the Web. The system converts documents (or the discrete files of a compound document) into HTML on the fly.

①FileNet's Web Publisher is a toolkit for the Panagon IDM (Integrated Document Management) system and can be used with imaging, COLD and workflow. Features include the PWP Project Manager, which draws source documents from Panagon libraries, organizes them into related sections and subsections and automatically publishes linked Web sites and online compound documents. Web sites can be published and managed from the Panagon libraries within the document repository.

PWP Station uses translation templates that control the automatic creation of HTML renditions of source documents, including format and style. The PWP Station also generates hyperlinks, tables of contents, keyword indexes and reference lists.

PWP Scheduler automates updates to Web publications on a set schedule or whenever new versions of existing materials become available.

Related project files maintain an audit trail of what was published at any given time, an important feature as the legal liabilities of Web publishing emerge.

①FileNet says it is waiting for enabling technologies, such as wider Web browser support, before incorporating XML. When this happens, the system will support the publishing of discrete content as well as compound documents in XML as well as HTML pages. Panagon Web Publisher is base priced at \$19,500 for the server and three clients with Web publishing stations.

①Documentum (Pleasanton, CA 925-4636800 [www.documentum.com](http://www.documentum.com)) describes "content" as any business-critical document or information in any format, such as word processing files, spreadsheets, graphics, emails, HTML, Java, etc.

RightSite is the module included in ①Documentum's Enterprise Document Management System (EDMS) that enables the assembly and delivery of content-based information to Intranets. Using RightSite, EDMS users can store, manage, retrieve and deliver mixed "content" from a common repository. All objects are stored in the DocuBase object-relational database and are shared with the RightSite Web server.

RightSite is not new, but support for XML has been added through integration with other tools for enhanced content management capabilities. ①Documentum promises to add native XML functionality in a future release. Presently RightSite supports XML through their partnerships with Abortext (Waltham, MA 781-529-1000 [www.arbortext.com](http://www.arbortext.com)), makers of Adept XML editor.

RightSite includes native support for SGML, which can be used along with HTML to bring content to the Web. WebQL, a super set of the Structured Query Language (SQL), is a published set of HTML directives for dynamic page assembly. WebQL is used to tag items with attributes and create live links to other items. RightSite automates the process of link creation and deletion.

Saillant Consulting Group's (Denver, CO 303-846-3088) HTMLRender is integrated with ①Documentum's RightSite to automate the conversion of documents to HTML to simplify creating Web-ready documents.

①Documentum says new APIs and SDKs will be offered in future releases to add more flexibility to RightSite as well as added support for active server pages and XML publishing. RightSite is included in F.DMS 98 and is not sold separately.

PC Docs (Burlington, MA 781-2733800 [www.pcdocs.com](http://www.pcdocs.com)) offers what they call a "compound document management and publishing module," called Docs Binder, for their Docs Open document management system. Docs Binder is geared toward collaborative environments where documents are made up of multiple items and have multiple authors spread across the enterprise.

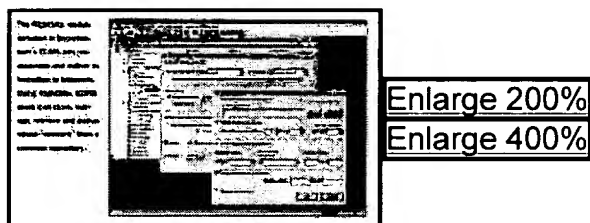
Within a "binder," the separate components of varied file formats are managed as individual documents. The documents can be edited separately or simultaneously within the larger binder document.

Individual documents can be used in more than one binder, allowing for the reuse of content. The content can be updated within all binders automatically.

The contents of the binder are viewed in an Explorer-like tree diagram. Documents can be dragged and dropped between and reordered within binders. Binders are stored in XML format and contain tags that identity links between documents.

For publishing, PC Docs integrates with third-party products such as PDFfusion from Computerized Document Control (Monmouthshire, UK 888-240-1752 [www.docctrl.com](http://www.docctrl.com)).

PDFfusion will publish binders as PDF documents on the fly. The software offers templates for layout. Other third-part publishing systems can be integrated using the Dots Binder toolkit.



The RightSite module included in Documentum's EDMS lets you assemble and deliver information to Intranets. Using RightSite, EDMS users can store, manage, retrieve and deliver mixed "content" from a common repository.

---

[^ Back to Top](#)[« Back to Results](#)[< Previous](#) Article 14 of 206 [Next >](#)[Publisher Information](#)[Print](#)[Email](#)☐ Mark Article[Citation](#) , [Full Text](#) , [Text+Graphics](#) , [Page Image - PDF](#)

Copyright © 2004 ProQuest Information and Learning Company. All rights reserved. [Terms and Conditions](#)

[Text-only interface](#)

From: ProQuest  
COMPANY